Modeling time to death for retirees in Kenya

BY

Margaret Ndisha Mwakala (B.Sc)

School of Mathematics

University of Nairobi

A dissertation submitted in partial fulfillment of the requirements for the award of the degree of Master of Science (Biometry) in the School mathematics of University of Nairobi

UNIVERSITY of Nairobi

July, 2013

Declaration

I declare that this research project is my original work and has not been presented for a degree in any other university or for any other award.

Margaret Ndisha Mwakala

----------------------- ------------------------

Signature Date

This research project is submitted for examination with my approval as university supervisor.

Dr. Nelson Owuor

School of Mathematics,

Mathematics department, ----------------------- ------------------------

University of Nairobi, Signature Date

Dedications

I here dedicate this project to the older generation of retirees, individuals who have gained experience and from whom we benefit a wealth of knowledge. I also dedicate this project, especially, in a special way, to my mother, for her unwavering strength and prayers for me and to my siblings and my friends. Specifically, to my boss, for allowing me to have time to do this project and for supporting me through out. For always reviewing my work, pushing me to be better, may God bless you.

Acknowledgement

I would like to acknowledgement my God, for being my source of strength to do this project, the insurance companies and pension schemes that provided me with the data, to make this project a success.

I would also like to acknowledge my supervisor, Dr. Nelson Owuor, for taking time off his busy schedule to work with me on this project and my classmates, for being a source of encouragement and for their continual support, my lecturers, supervisors and colleagues.

Keywords

Retirement: is the point where a person stops employment completely

Pension: is a contract for a fixed sum to be paid regularly, in regular installments, to a person, typically following retirement from service.

Pension schemes: A tax-privileged individual investment vehicle, with the primary purpose of building a capital sum to provide retirement benefits, although it will usually also provide death benefits.

Life Annuity: is a financial contract in the form of an insurance product according to which a seller (issuer) — typically a financial institution such as a life insurance company — makes a series of future payments to a buyer (annuitant) in exchange for the immediate payment of a lump sum (single-payment annuity) or a series of regular payments (regular-payment annuity), prior to the onset of the annuity.

Insurance: is the equitable transfer of the risk of a loss, from one entity to another in exchange for payment. It is a form of risk management primarily used to hedge against the risk of a contingent, uncertain loss.

Life expectancy: is the expected (in the statistical sense) number of years of life remaining at a given age. It is denoted by e_x which means the average number of subsequent years of life for someone now aged, X, in our case, average of 55, according to a particular mortality experience.

Life tables: is a table which shows, for each age, what the probability is that a person of that age will die before his or her next birthday

Abstract

Many countries have studied the life of retirees, with special attention given to the quality of life of these individuals. In Kenya, no work has been done on the determinants of time to death after retirement. The average age of survival after retirement in Kenya is also not known. It is necessary to determine the causes, if any, of mortality in the retirement years and necessary interventions put in place to address them. The life expectancy of a retiree in Kenya is also a key matter, especially in life annuity programs in the insurance companies in Kenya, with the mortality and longevity risks being of great interest to actuaries and financial professionals. This study attempts to determine the life expectancy of a retiree in Kenya, both for men and women, by using the life table methods, widely used to determine the life expectancy. This information will be very useful in determining the cessation of life annuities, used usually to determine the end of guarantee period, which is calculated with the life expectancy in mind.

As we determine the life expectancy, the determinants of time to death is also key, to shed light on what makes one retiree live longer compared to another retiree. Information like this can assist employers to advise their employees well on better planning, to ensure that retirement doesn't lead to death, as has been seen over the years [3].

Previous work on time to death after retirement has been done in other countries, specifically, developed countries, [3]. Most studies associated early retirement, by extension age at retirement, gender and the socio-economic status of the retirees' with mortality. To date no consensus has been reached on the survival or mortality of people who retire early compared to those who retire later. Some researchers concluded that early retirement harms health, attributing this to illness before retirement or the change of life events associated with retirement. On the other hand, there is a widespread perception that early retirement is associated with longer life expectancy and that retiring later leads to early death. The possible health benefits of retirement, such as reduced role demand and a more relaxed lifestyle, have been postulated to improve longevity among people who retire early [4].

While the average years lived has been seen to generally increase in both developed and developing countries over time, determinants to death after retirement can hardly be assumed to be the same in the different set-ups, given the resources and medical provisions in the different

set-ups [4]. This study, in view of the finding of other studies done, attempts to check the validity of the models used elsewhere to determine time and the determinants to death in the Kenyan set-up, taking into consideration the age at retirement, socio-economic status, derived from the monthly pension paid to the retiree and gender, the time to death will be determined, thus observing the impact of certain factors as will be described below.

The cox proportional hazard regression model has been used widely to determine survival time, taking into considerations its strength of considering individuals who had the event of interest (death) and those who were censored [9]. In this study, we apply this model, to further explain the determinants of time to death after retirement in Kenya, i.e. sex, age-at-retirement and socio-economic status.

This study showed that there is a significant difference between the survival times between male and female retirees, in fact showing that male retirees are 2.1 times more likely to die, compared to their female counterparts. This therefore leading us to conclude that gender does indeed determine the rate of survival after retirement in Kenya. Socio-economic status and age-at-retirement were also found to be significant determinants to time to death after retirement.

This study concluded that early retirement does indeed increase the chances of early death, though the study also suggests that more information on the health status of the early retirees should be used as a covariate during such a study, to eliminate the possibility that the early retirees were already sick, thus their health status rather that their age cause the early death. Retirees should be advised to plan for their retirement since those retirees who earned less were 25% more likely to die than their counterparts who earned more. The standards of living after retirement can be boosted by a better savings plan that ensures that the life after retirement is catered for therefore employers should educate their employees on planning for this.

Table of contents

List of tables and figures

1.0 Introduction

Over the past decades, there has been a consistent decrease in the mortality rates in most developed countries. Recently, developing countries, including Kenya, have been experiencing a similar trend, thereby implying that people are living longer years, even after retirement. According to Kenya National Bureau of Statistics, life expectancy which was estimated at 48.9 years in 2006 has risen to 64 years in 2012 [13]. While this has been observed, this study seeks to quantify the life expectancy after retirement, to determine the average years lived after retirement by men and by women in Kenya.

The process of retirement begins with attitudes toward retirement, retirement policies, and factors in the decision to retire. Currently, high levels of desire for retirement and poor health cause most retirements to occur at or near the minimum age for retirement. The retirement transition has varying effects, depending on how the individual arrives at retirement. Those who retire voluntarily have little or no difficulty adjusting [10]. Those who are forced by mandatory retirement policies tend to be dissatisfied at first, but eventually they adjust. And those who retire because of poor health are understandably the most dissatisfied, although retirement improves health for many of them. Retirement itself has no predictable negative effect on physical health, self esteem, or life satisfaction. It does tend to reduce activity level, therefore, since the most precious, creative and innovative period in one's life is the 10-year period around the age of 32, one should plan their career path to use this precious 10-year period wisely and effectively to produce the greatest achievements in their life, [1]. Work is a determinant of identity, self actualization; self worth etc. Studies have shown that inactivity in retirement increases the death rates of retirees. However, when you get older, you should plan your career path and financial matter so that you can retire comfortably at the age of 60 or earlier to enjoy your long, happy and leisure retirement life into your golden age of 80s and beyond, [1]. In retirement, you can still enjoy some fun work of great interest to you and of great values to the society and the community, but at a part-time leisure pace on your own term.

More studies have been done to support the concept of early retirement, alluding to the fact that early retirees tend to die earlier, [3], others have found that there was no significant difference on mortality between the early retirees and those who worked up to the ages of 65 and above, [2].

While studies on life after retirement shows that death in old age is associated with age at retirement, the likely determinants of death in retirees have been done in developed countries with no studies conducted in developing countries.. This is due to the fact that unbiased population data are not readily available for many countries, especially developing countries. For our - retrospective- cohort analysis, we were able to make use of a data set from an insurance annuity program and a pension scheme for civil servants. The data accounted for the age, monthly pension paid, the age at retirement and gender. Most of the pension schemes do not take into account variables such as the health status, marital status and the profession of retirees which would have shed more light on the reasons for retirement that would explain mortality rates among these individuals.

Despite the challenges with the data, this project seeks to address the issue of mortality after retirement in Kenya. We seek to assess the determinants to mortality and determine the life expectancy of individuals after retirement in Kenya.

1.1 Background

Retirement is the point where a person stops employment completely (Wikipedia, http://wikipedia.org/wiki/retirement, February 2013). In Kenya, the retirement age has been 55 years for a long time but since 2009, this age has been changed to 60 years. In a statement to the press, the then Public Service Minister, Dalmas Otieno, also announced a new contributory public service scheme and said that the age raise would give the government time to programme for resources to fund the new pension scheme and allow room for introduction and operationalization of the (public service) scheme. This however didn't affect the early retirement age, which remained at 50 years and other areas like retirement in public interest, medical grounds, from the disciplined services or due to abolition of offices will remain the same. The new contributory public service scheme, which also kicks off this month (July 2013), will require civil servants to make a monthly contribution of 7.5 percent from their salaries while the government will put in 15 percent. Public service employees will also be able to access their contributions upon leaving employment and only wait for the government's contribution at age 65. The Minister further said that deductions will be increased over three years before hitting the 7.5 percent of pay, [14].

Concerns were raised on this as many Kenyan's felt that this would deny the youth from employment because essentially, retirement creates employment for the younger generation. Most of the Kenyans and the government as well are more concerned about the youth than the retirees, which is to be expected, why? Because people don't wake up one day and it's time to retire, people are well aware that retirement is coming, 20 or even 30 years to come. Which raises questions, do Kenyans prepare for retirement? What happens to people after they retire? Are the myths about death after 2 to 3 years true? What would be the cause of these deaths?

Retirement is a passage from one lifestyle to another. One way to think of the term "retire" is by placing a hyphen between the 'e' and the 't' and creating a new term---re-tire: To put on new treads.

Those who take the voyage seriously and do the right kind of retirement planning usually have a smoother trip and more fun. It has been noted that some people spend more time planning a two-week vacation than they do retirement planning. [15]

With these questions and myths in mind, Kenya needs to find out the state of its retirees and assess if enough is being done by employers and the government itself to prepare employees for retirement. Is a 2-week class on the retirement month enough? Should we start this process early? Who is responsible to disseminate the education to the young employees, leaving and spending merrily oblivious of the yester years?

1.2 Objectives

1.2.1 Main objective of the study

To investigate the determinants of time to death of retirees in Kenya

1.2.1 Specific objectives of the study

Analyze the data and determine:

1. If gender, socio-economic status and age-at-retirement are significant determinants of time to death in retired Kenyans.
2. If there is a significant difference in time to death between male and female retirees and the different pension categories, (What is the strata in this case) in the study population

3. The life expectancy, e, after retirement in Kenya. What is the average time to death after retirement in Kenya?

1.2.2 Research questions

1. Do gender, socio-economic status and age-at-retirement significantly determine time to death in retired Kenyans
2. Is there a significant difference in time to death in the strata in the study population
3. What is the average time to death after retirement in Kenya?

1.3 Hypothesis

There is no significant difference between the strata in the study population, that is, there is no difference in time to death between males and females or among the different categories of monthly pension received, thus the hypothesis:

$$\boldsymbol{H_o}: \boldsymbol{S_{(a)}} = \boldsymbol{S_{(b)}}$$

Versus

$$\boldsymbol{H_1}: \boldsymbol{S_{(a)}} \neq \boldsymbol{S_{(b)}}$$

Where $\boldsymbol{S_{(a)}}$ is the survival function for the males while the $\boldsymbol{S_{(b)}}$ is the survival function for the females.

There are no specific causes to the deaths of retirees in Kenya. This then means that given the cox proportional model used for analysis of the determinants of the hazards (risk of outcome, in my case death) are not significant, with their coefficients being equal to 0, thus assuming:

$$\boldsymbol{H_o}: \boldsymbol{\beta_l} = \boldsymbol{0}$$

Versus

$$\boldsymbol{H_1}: \boldsymbol{\beta_l} \neq \boldsymbol{0}$$

Where *l* is the number of covariate considered in the model, in my case, 3, that is:

1. Age at retirement
2. Category of the monthly pension received
3. Gender

1.4 Justification

Studies have been done on the impact of early retirement, with results showing that self perceived health, employment grade and job satisfaction are all independent predictors of early retirement [12]. In Kenya, early retirement was hugely embraced, but the impact of the same has not been assessed. The retirement age was also changed to 60, adding 5 years to the normal retirement age, a change whose impact hasn't also been assessed, therefore, the average age of survival after retirement in Kenya is not known. The quality of life of the retirees is also a hugely ignored matter. This study attempts to determine the determinants to time to death, if any, of mortality in the retirement years and necessary interventions put in place to ensure that these causes are addressed. Since studies have been done in other countries, it is interesting to check the validity of the models used elsewhere to determine time and the determinants to death in the Kenyan set-up.

2.0 Literature review

There is a widespread perception that early retirement is associated with longer life expectancy and later retirement is associated with early death. No consensus has been reached on the comparative survival or mortality of people who retire early or late. In the paper, [4], after analyzing data from past employees of Shell, using the Kaplan Meier curves and cox proportional model (used wald X2 test to test significance of the hazard ratios), on comparison of survival curves on different retirement ages, found that age at retirement is not associated with increased or decreased survival. However, result showed that gender and socio-economic status of the employee was a determinant of survival after retirement.

In one of the most well-done (statistically speaking) studies of retirement and health, researchers found that retiring had a negative impact on health. This study looked at 16,827 men in Greece who had not been diagnosed with a health condition, such as diabetes, stroke, cancer or heart disease. It compared men who retired to men who were still working (remember, none of them had a major health condition at the beginning of the study). The study discovered that the retirees had a 51% increase in their risk of death as compared to those still working.(after controlling for variables, such as wealth, education, marital status, etc.). Most of the increase in death risk was linked to heart disease and cardiovascular health, [5]. It would be important to assess the quality of life of the retirees. For example, with the wealth of information gathered as at 50 years, is it wise to let go this individual or should there be alumni programs where the experience gained over decades is transferred to the younger generation? Programs like this would benefit both the youth and the older folk, improving greatly their quality of life.

Hilke Brockmanna, [13], used Kaplan Meier curves to determine if there was a difference in the survival rates of different retirement ages, socio-economic status (those who retired due to reduced earning capacity) and used the cox proportional models to determine the determinants to death after retirement. Results stated that early retirement impacts long-term survival significantly. Lower survival chances were observed for persons with poor health at early retirement compared with pensioners who retire later. However, employees who leave the labor market early and healthy may have better survival chances than people retiring between the ages of 61 and 65. Comparisons of deaths (cases) within 5 years of retirement with survivors (controls) showed that pre-retirement health status was the only significant predictor of survival after early retirement. Among normal retirees, lower status workers were more likely to die within 3 years of retirement than higher status workers, who were more prominent among deaths 4 to 5 years after retirement [11].

A study conducted by Sanlam Employees Benefit (in South Africa) shows that people are not saving enough for retirement. Mr Dawie De Villiers of Sanlam's Structured Solution in his presentation mentioned that of the surveyed pensioners only 40% believed that they have saved enough capital to last for the rest of their lives. Pensioners also highlighted that if there was anything they would have done differently is that they would have planned better and saved more before retiring. This supports the findings of the study, “people are not saving enough for retirement”. Most will have to work after they reach their retirement age, not because they want to , but because they need funds, or live with their children or rely on the social old age pension.

Results presented show that this area of study requires more studies to be done to determine the determinants of longer or shorter survival years, after retirement.

2.1 Causes of death in old age

Most studies associate death after retirement to:

1. Health problems -- Illness and disability; chronic or severe pain; cognitive decline; damage to body image due to surgery or disease.
2. Loneliness and isolation -- Living alone; a dwindling social circle due to deaths or relocation; decreased mobility due to illness or loss of driving privileges.
3. Reduced sense of purpose -- Feelings of purposelessness or loss of identity due to retirement or physical limitations on activities.
4. Fears -- Fear of death or dying; anxiety over financial problems or health issues.
5. Recent bereavement -- The death of friends, family members, and pets; the loss of a spouse or partner.

However, it is important to mention that although retirement can be painful, but it is not lethal. It is a change not unlike those earlier in life. While these causes are duly documented, data on the same is unavailable, since these causes would allude to the quality of life of the retiree, which would have direct impact on the longevity of their life. Also, most people who retire early or die shortly after retirement probably had health problems before they stopped working. Retirement had nothing to do with their demise. This brings into question the health status of the retirees unfortunately; most pension schemes in Kenya don't ask about the health status of the retiree thus it's impossible to make inference on this as a cause of early death.

However, the age-at-retirement will be a variable to shed some light on if it's a significant determinant cause of death. In his article, Michael E. Leonetti mentions that the only connection between retirement and early death may be that some retirees fail to keep active. They relax to the point that their bodies self-destruct. They give up. They fail to stay in charge. There are many reasons for retiring early, and there are just as many for retiring later. But avoiding retirement planning and staying on the treadmill because you fear retirement will cause early death and this should not be the case.

2.2 Survival analysis

Survival analysis is just another name for time to event analysis. The term survival analysis is used predominately in biomedical sciences where the interest is in observing time to death either of patients or of laboratory animals. Time to event analysis has also been used widely in the social sciences where interest is on analyzing time to events such as job changes, marriage, birth of children and so forth. The engineering sciences have also contributed to the development of survival analysis which is called "reliability analysis" or "failure time analysis" in this field, since the main focus is in modeling the time it takes for machines or electronic components to break down.

Survival analysis techniques allow for a study to start without all experimental units enrolled and to end before all experimental units have experienced an event. This is extremely important because even in the most well developed studies, there will be subjects who choose to quit participating, who move too far away to follow, who will die from some unrelated event, or will simply not have an event before the end of the observation period. The researcher is no longer forced to withdraw the experimental unit and all associated data from the study. Instead, censoring techniques enable researchers to analyze incomplete data due to delayed entry or withdrawal from the study. This is important in allowing each experimental unit to contribute all of the information possible to the model for the amount of time the researcher is able to observe the unit.

There are certain aspects of survival analysis data, such as censoring, where a censored observation is defined as an observation with incomplete information, and non-normality, which generate great difficulty when trying to analyze the data using traditional statistical models such as multiple linear regressions. The non-normality aspect of the data violates the normality assumption of most commonly used statistical model such as regression or ANOVA. The point of survival analysis is to follow subjects over time and observe at which point in time they experience the event of interest. It often happens that the study does not span enough time in order to observe the event for all the subjects in the study. This could be due to a number of reasons. Perhaps subjects drop out of the study for reasons unrelated to the study (i.e. patients moving to another area and leaving no forwarding address). The common feature of all of these examples is that if the subject had been able to stay in the study then it would have been possible to observe the time of the event eventually.

The other important concept in survival analysis is the hazard rate. From looking at data with discrete time (time measured in large intervals such as month, years or even decades) we can get an intuitive idea of the hazard rate. For discrete time the hazard rate is the probability that an individual will experience an event at time t while that individual is at risk for having an event. Thus, the hazard rate is really just the unobserved rate at which events occur. If the hazard rate is constant over time and it was equal to 1.5, for example, this would mean that one would expect 1.5 events to occur in a time interval that is one unit long. Furthermore, if a person had a hazard rate of 1.2 at time t and a second person had a hazard rate of 2.4 at time t then it would be correct to say that the second person's risk of an event would be two times greater at time t. It is important to realize that the hazard rate is an un-observed variable yet it controls both the occurrence and the timing of the events. It is the fundamental dependent variable in survival analysis.

Another important aspect of the hazard function is to understand how the shape of the hazard function will influence the other variables of interest such as the survival function. The hazard function may not seem like an exciting variable to model but other indicators of interest, such as the survival function, are derived from the hazard rate. Once we have modeled the hazard rate we can easily obtain these other functions of interest. To summarize, it is important to understand the concept of the hazard function and to understand the shape of the hazard function.

3.0 Methodology

This study was done using data from different pension schemes and insurance annuity programs in Kenya. Data was cleaned, survival dataset created and exploratory and inference analyses done to explore the determinant of time to death in retirees in Kenya.

3.1 Study area

The study was conducted in Kenya, a country located in East Africa bordered by Uganda, Tanzania, Somalia, Ethiopia and Sudan. The study population included retired staff that were on a pension scheme or annuity program with certain insurance schemes. This included people who received lump-sum after retirement and joined a scheme that paid them a calculated pension amount either annually, semi-annually, quarterly or monthly. This gave us information to make it possible to follow an individual from pension payment commencement date to the date of death (if any) or consider the individual censored.

3.2 Data description

The data contained 5397 observations and 349 failures (individuals who experienced the event) in the data used for this analysis.

The dependent variable considered was survival time since retirement and the covariates: sex, age-at-retirement and categories of pension received monthly. Of the covariates, sex and pension category were the categorical variable and age at retirement was a continuous variable.

3.2.1 Exploratory data analysis

Exploratory data analysis, to further describe the data, to understand the kind of individuals in the dataset used for this project will be done, based on:

1. Gender
2. Pension categories
3. Age categories

3.2.2 Data preparation

Data was sourced from various insurance companies and pension schemes in Kenya. Data available had variables that were used to create survival data for this project.

We were keen on the variables:

1. Date of birth (time of origin for all individuals in the data)
2. Date of retirement
3. Date of death (if death occurred, the date is provided, if not, the end of study date is indicated, i.e. **31st March 2013**)
4. Status , censored or died
5. Gender
6. Monthly pension being received

The variable age at retirement was determined by finding the difference between "date of birth" and "date of retirement".

The monthly pension variable has values with great disparities, for this reason, we decided to create categories for the amounts, having:

<5000 coded as 1

>5000 and <=10000 coded as 2

>10000 and <=20000 coded as 3

>20000 and <=40000 coded as 4

>40000 and <=100000 coded as 5

>100000 coded as 6

Finally, we had six categories to represent the socio-economic status of the retirees.

A survival dataset was created in stata, with the variables:

1. **_st** variable is a 0/1 variable, equal to 1 for observations whose data has been stset (it would be zero if one had excluded some cases with an, if qualifier, for instance).
2. The **_d** variable is the censoring indicator, another 0/1 variable, and corresponds to the variable died in this case.
3. The variable **_t** is the duration variable, corresponding to the difference between the date of retirement and the date of death (for those retirees who died), with our time scale being years
4. **_t0** is a variable recording the date of entry to the study for each case, i.e. when they were first observed at risk of the event. For this dataset, it's the retirees' date of retirement, which is different for most of them.

3.3 Data Analyses

In this study Kaplan Meier curves were used to provide insight into the shape of the survival function for each group and give an idea of proportionality. In any data analysis it is always a great idea to do some univariate analysis before proceeding to more complicated models. In survival analysis it is highly recommended to look at the Kaplan-Meier curves for all the categorical predictors. This will provide insight into the shape of the survival function for each group and give an idea of whether or not the groups are proportional (i.e. the survival functions are approximately parallel). We also consider the tests of equality across strata to explore whether or not to include the predictor in the final model. For the categorical variables we will use the log-rank test of equality across a stratum which is a non-parametric test. For the continuous variables we will use a univariate Cox proportional hazard regression which is a semi-parametric model.

3.4 Life table analysis

Life tables refer to a statistical procedure which generates duration (time to event) distributions for an entire dataset or separately for each level of a factor. As such it is a form of nonparametric survival analysis. The primary output is a life table in which the rows are researcher-defined time intervals and the columns have to do with counts, probabilities, and cumulative probabilities of the event of interest occurring during the given time interval. The Life Table Method was identified as the most suitable approach to analyze the pattern of mortality for this data set. Life Table analysis provides estimates of probabilities of surviving a given number of years after retirement. This technique allows subjects to enter (i.e., retire at different ages) or leave the study (i.e., die) at different points in time and it utilizes all the data on partial exposure to the risk of dying. It is non-parametric and requires no assumptions about the distribution of the survival function. The table captures:

x - Age or duration; $0 \le x \le \omega$

l(x) - number of persons living at age x

d(x) - number of deaths between age x and x+1

q(x) - is the probability of dying between age x and x+1

p(x) - is the probability of surviving between age x and x+1

L(x) - person years lived between age x and x+1

T(x) - Total persons-years lived after age x

e(x) - life expectation at age x, (T(x)/l(x)

3.5 Theory of survival distribution functions

The variable of interest (**response variable**) was the time to death after retirement or censoring time, with the study ending on the 31st March 2013 meaning that all retirees who didn't experience the event of interest, death after retirement, were considered censored, with a censoring variable, *Status*, which indicated whether a retiree was censored or died, as another key variable.

Hence let

$\boldsymbol{Y_j}$ Be the time to death after retirement or censoring time for the **j**th retiree.

$\boldsymbol{\delta_j}$ Be the censoring variable, *Status,* which indicates whether the **j**th retiree was censored or died

$\boldsymbol{T_j}$ Be the time to death for the **j**th retiree.

$\boldsymbol{C_j}$ Be the censoring time for the **j**th retiree.

$$\boldsymbol{Y_j} = \begin{cases} T_j\,, & \delta_j = 0 \\ C_j\,, & \delta_j = 1 \end{cases} \tag{1.1}$$

$\boldsymbol{Y_j \geq 0,}$

$\boldsymbol{T_j \geq 0}$

$\boldsymbol{C_j \geq 0}$

Hence in the dataset there were **r** happenings of the event of interest (**death after retirement)**.

Thus let

$\boldsymbol{t_i}$, be the time of death, for the **i**th death. Where ***i=1, 2, 3….r***

To achieve the objective of comparing the survival distribution functions for both males and females, there was need to estimate their respective survival functions using the Kaplan Meier estimator with the assumption that:

1. Probabilities for the event of interest will depend only on time after the initial event--they are assumed to be stable with respect to absolute time. That is, cases that enter the study at different times (for example, retirees who retired at different times) should behave similarly.

2. There should also be no systematic differences between censored and uncensored cases. If, for example, many of the censored cases are patients with more serious conditions, then the results may be biased.

$\boldsymbol{S}(\boldsymbol{t}) = \mathbf{Pr}(\boldsymbol{T} > t),$ (1.2)

the probability of survival beyond time **t**, i.e. the time to event (**T**) occurring after time **t**.

Hence let

$\boldsymbol{n_i}$, be the number of retirees at risk of death at time $\boldsymbol{t_i}$, representing all retirees who have not died, including those who may be censored after time $\boldsymbol{t_i}$

$\boldsymbol{d_i}$, be the number of deaths observed at time $\boldsymbol{t_i}$.

By description the conditional probability of surviving past time $\boldsymbol{t_i}$given survival to that time is estimated by $\frac{(\boldsymbol{n_i} - \boldsymbol{d_i})}{\boldsymbol{n_i}}$, or simply: $\mathbf{1} - \frac{\boldsymbol{d_i}}{\boldsymbol{n_i}}$ (1.3)

Overall the unconditional probability of surviving beyond any time **t** is estimated by the equation below;

$$\widehat{\boldsymbol{S}}(\boldsymbol{t}) = \prod_{\boldsymbol{t_i} \leq \boldsymbol{t}} \frac{(\boldsymbol{n_i} - \boldsymbol{d_i})}{\boldsymbol{n_i}} \quad (1.4)$$

$\prod$: *represents the product of probalility of survival through out the observed time, t*

The equation above is known as the **Kaplan-Meier estimator which can be constructed using statistical software**. The generation of the estimator is as indicated in the table below:

Table 1 - Kaplain-Meier Estimator generation

i	t_i	d_i	c_i	n_i	$(n_i - d_i)$	$\frac{(n_i - d_i)}{n_i}$	$\hat{S}(t)$
0	t_0	d_0	c_0	n_0	$(n_0 - d_0)$	$\frac{(n_0 - d_0)}{n_0}$	$\frac{(n_0 - d_0)}{n_0}$
1	t_1	d_1	c_1	n_1	$(n_1 - d_1)$	$\frac{(n_1 - d_1)}{n_1}$	$\frac{(n_0-d_0)}{n_0} * \frac{(n_1-d_1)}{n_1}$
.							
k	t_k	d_k	c_k	n_k	$(n_k - d_k)$	$\frac{(n_k - d_k)}{n_k}$	$\frac{(n_0-d_0)}{n_0} * \frac{(n_1-d_1)}{n_1} * \dots * \frac{(n_k-d_k)}{n_k}$

3.6 Inference – Using Log Rank tests

3.6.1 Comparing Survival Curves

An interesting issue in the clinical practice is the comparison of the survival functions of two (or more) groups of subjects. Classical statistical methods such as T-test or Chi-square Test are not appropriate to compare the proportion of survivors for the reasons described before. Alternatively, one could compare the proportions of survival at one or more pre-specified time points, but this approach ignores the total survival experience of the groups during the whole follow-up period. Moreover, the time points are arbitrary chosen. The most common method used to compare survival curves is then a statistical hypothesis test called log-rank test [6], which takes the whole follow-up period into account and does not require any assumption about the distribution of the survival function, which is rarely normally distributed and is often skewed because it often comprises many early events and only a few late ones. The null hypothesis for the log-rank test is that there is no difference between the survivals of two or more populations that are being compared (i.e. the probability of the event of interest occurring at any time point is the same for each population). The comparison is based on the difference between the observed number of events in each group and the expected number of events in case of non-difference between the two groups.

To compare whether the survival distribution functions for the male and female retirees are statistically significantly different, there is need to test the hypothesis below. We would have to test the hypothesis below:

$\boldsymbol{H_o}: \boldsymbol{S_{(a)}} = \boldsymbol{S_{(b)}}$

Versus

$\boldsymbol{H_1}: \boldsymbol{S_{(a)}} \neq \boldsymbol{S_{(b)}}$

Where $\boldsymbol{S_{(a)}}$ is the survival function for male retirees while the $\boldsymbol{S_{(b)}}$ is the survival function for the female retirees.

Hence let:

$\boldsymbol{n_{mi}}$, be the number of retirees at risk of dying at time $\boldsymbol{t_i}$**,** for the **m**$^{\text{th}}$ group. (2.1)

$\boldsymbol{d_{mi}}$, be the number of dead retirees, whose death was observed at time $\boldsymbol{t_i}$**,**

for the **m**$^{\text{th}}$ group. (2.2)

Let **X** be the random variable that takes the values, $\boldsymbol{d_{mi}}$**.**

Recall $\boldsymbol{m = 1, 2}$**;** there are two groups, strata, male and female retirees

A log-rank test needs 2* 2 contingency table with the variables described in the slide before to obtain the test statistic.

Table 2-Contingency table to describe log-rank test for 2 strata (Male and Female)

Group	No. dead	No. Alive	No. At Risk
Males	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Females	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
Total	d_i	$n_i - d_i$	n_i

The log-rank test statistic follows a hyper-geometric distribution hence the probability of the random variable **X** taking the value d_{mi}, would be given by

$$Prob(X = d_{mi}) = \frac{\binom{n_{1i}}{d_{1i}}\binom{n_{2i}}{d_{2i}}}{\binom{n_i}{d_i}} \qquad (2.3)$$

While the expectation and variance of the random variable **X** are given as below

$$E(X = d_{1i}) = \frac{d_i n_{1i}}{n_i} \text{ and } Var(X = d_{1i}) = \frac{n_{1i} n_{2i} d_i (n_i - d_i)}{{n_i}^2 (n_i - 1)} \qquad (2.4)$$

Let **Y**, the total number of dead male retirees.

$$Y = \sum_{t_i} d_{1i} \qquad (2.5)$$

Then

$$E(Y) = \sum_{t_i} E(d_{1i}) \text{ and } Var(Y) = \sum_{t_i} Var(d_{1i}) \qquad (2.6)$$

Let **Z** be defined as below, after standardization of the variable Y,

$$Z = \frac{Y - E(Y)}{\sqrt{Var(Y)}} \qquad (2.7)$$

$$E(Y) = 0 \text{ and } Var(Y) = 1 \text{ consequently } Z \sim N(0,1). \qquad (2.8)$$

Substituting the values of Y, E(Y) and Var(Y), we have Z defined as:

$$Z = \frac{\sum_{t_i} d_{1i} - \sum_{t_i} E(d_{1i})}{\sqrt{\sum_{t_i} Var(d_{1i})}} \qquad (2.9)$$

Then from probability distribution if consequently $Z \sim N(0,1)$ then $Z^2 \sim \times^2\ (1)$.

Thus Z^2 is known as the log-rank test statistic and it follows a chi-square distribution with one degree of freedom.

$$Z^2 = \frac{\left\{\sum_{t_i}\left[d_{1i} - \frac{d_i n_{1i}}{n_i}\right]\right\}^2}{\sum_{t_i} \frac{n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}} \qquad (2.10)$$

Thus to conduct the hypothesis test for the hypotheses below:

$H_o: S_{(a)} = S_{(b)}$

Versus

$H_1: S_{(a)} \neq S_{(b)}$

You obtain a critical value from the chi-square distribution tables at ∝ level of significance. If the **Test Statistic (Z^2) is greater than the Critical Value** then reject $H_o: S_{(a)} = S_{(b)}$.

3.7 Inference – using Cox-Proportional Regression Model

The previous survival models described above shows that survival data can be analyzed using univariate approach therefore to compute a multivariate survival analysis, cox model was used[6].

3.7.1 Cox-Proportional Regression Model

The cox proportional hazard regression model extends the methods of non-parametric Kaplan-Meier estimates to regression type arguments taking into account the effects of censored observations [7]. The principle of Cox proportional hazards model is to link the survival time of an individual to covariates. For example, in the medical domain, when finding out which covariate has the most important impact on the survival time of an individual, Cox advances to predict survival time in individual subjects by only utilizing variables covarying with survival and ignoring the baseline hazard of the individuals. The model uses the assumption that the hazard functions of different individuals remained proportional and constant over time.

When there are several continuous explanatory variables, it is much more useful to use a regression method such as Cox rather than a KM .

Assumptions of the Cox proportional hazards regression model:

1. Non-informative censoring. To satisfy this assumption, the design of the underlying study must ensure that the mechanisms giving rise to censoring of individual subjects are not related to the probability of an event occurring.

2. That the hazard functions are proportional over time.

3. No ties in the event times.

Cox's proportional hazards model is a distribution free model in which predictors are related to lifetime multiplicatively.

To determine the determinants of time to death in the study, cox proportional hazards model regression analysis was computed to test whether the regression coefficients in the model pertaining to each of the predictor variables including the age-at-retirement are statistically significantly different from zero. The hypothesis tested was:

$\boldsymbol{H_o}: \boldsymbol{\beta_l} = \mathbf{0}$

Versus

$\boldsymbol{H_1}: \boldsymbol{\beta_l} \neq \mathbf{0}$

The cox proportional hazards regression model is the most popular approach to modeling the predictors to estimate their effects on time to event. It is semi-parametric model because it takes no assumption of the probability distribution function for the time to event; the baseline hazard takes any form and the predictors enter the model through the linear predictor, $\boldsymbol{n_l}$, . It is based on the assumption that the hazards are proportional.

The cox proportional hazards regression model is based on the hazard function, $\boldsymbol{h(t, X)}$at time $\boldsymbol{t}$

If we let $x_1, x_2, \ldots, x_p$be the values taken by the following covariates$X_1, X_2, \ldots, X_p$**,** where **p** is the total number of predictor variables in a model.

Then the linear predictor, $\boldsymbol{n_l}$, is given as

$$\boldsymbol{n_l} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (3.1)$$

Consequently the cox proportional hazards regression model is expressed as follows

$$\boldsymbol{h(t, X) = h_0(t) exp\left(\sum_{l=1}^{p} \beta_l X_l\right)} \quad (3.2)$$

$\boldsymbol{h_0(t)}$Is the baseline hazard function at time t.

β_lIs the **l**thregression parameter for the cox proportional hazards model, while $X_l$is **l**thpredictor variable.

Where

$$\boldsymbol{h_0(t)} \geq \mathbf{0}$$

and

$$\sum_{l=1}^{p} \beta_l X_l \geq \mathbf{0}$$

Recall $\boldsymbol{l} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{p}$**,** is the number of predictors in the cox proportional hazards model

3.7.2 Relative risk type measure type of association

The simple interpretation of the measures of association given by the Cox model as "relative risk or hazard ratio" is very desirable in explaining the risk of event for certain categories of covariates or exposures of interest. For example, when a two-level (dichotomous) covariate with a value of 0=no and 1=yes is observed, the hazard ratio becomes eβ where β is the parameter estimate from the regression. If the value of the coefficient β = 1.099, then e1.099 = 3. The measure is simply saying that the subjects labeled with 1 (yes) are three times more likely to have an event than the subjects labeled with 0 (no). In this way we have a measure of association that gives insight into the strength and direction of the relationship between our exposure and outcome [9].

One advantage of the cox proportional hazards model is that it takes care of censored observations aside from it being semi-parametric and ability to provide the relative risk measure for categorical covariates.

The proportionality of the hazards is an important assumption of the cox proportional hazards model, it means that the hazard ratio is constant over time, or more precisely the hazard for an individual observation is proportional to the hazard for any other individual observation.

The hazard ratio (HR) is obtained as shown from the ratio of hazard functions.

Let $\boldsymbol{h}(\boldsymbol{t},\boldsymbol{X}) = \boldsymbol{h_0}(\boldsymbol{t})\boldsymbol{exp}\left(\sum_{l=1}^{p}\beta_l X_l\right)$ be the hazard function for individual A (4.1)

And

Let $\boldsymbol{h}(\boldsymbol{t},\boldsymbol{X}^*) = \boldsymbol{h_0}(\boldsymbol{t})\boldsymbol{exp}\left(\sum_{l=1}^{p}\beta^*{}_l X^*{}_l\right)$ be the hazard function for individual B (4.2)

Then the hazard ratio (HR) is given by $\mathbf{HR}=\frac{\boldsymbol{h}(\boldsymbol{t},\boldsymbol{X})}{\boldsymbol{h}(\boldsymbol{t},\boldsymbol{X}^*)}=\exp\left[\sum_{l=1}^{p}(x_l{}^* - x_l)\right]$ (4.3)

Hence when the value above for the individuals A and B is constant then the proportional hazards assumption is satisfied.

3.7.3 Use of the partial likelihood function

The Cox model has the flexibility to include time-dependent explanatory variables and handle censoring of survival times due to its use of the partial likelihood function. The likelihood function for the proportional hazards model can be thought of in two parts. The individual hazard and the exponentiated function of the independent variables represented by the linear sum of the βixi's. The baseline hazard multiplied by the function of independent variables produces the hazard for the ith subject. The partial likelihood formula considers probabilities for those subjects who fail and does not explicitly consider probabilities for censored subjects. However, survival time information prior to censorship is used for those subjects who are censored. That is, the subject who is censored after the jth failure time is part of the risk set used to compute the jth likelihood even though this subject is censored later. We get estimates by finding values for the function of independent variables (betas) that maximize the partial likelihood. Some efficiency in estimate is lost but the model is robust and two of three standard properties of maximum likelihood estimates persist, being consistent and being asymptotically normal.

To estimate $\beta_l$the **lth**regression parameter for the cox proportional hazards model we'll use the partial likelihood method. We shall only estimate for **lth**covariate.

Steps to obtaining the Partial Likelihood

- Arrange the event times in order i.e. $t_1 < t_2 <, \dots, < t_k$. (5.1)
- Obtain the risk sets $R(t_1), R(t_2), \dots, R(t_k)$ (5.2)
- Then obtain the probability of an individual dying at time t_i given risk set $R(t_i)$, $t_i /$ $R(t_i)$, which is given by $= \frac{\boldsymbol{exp}(\beta X_i)}{\sum_{r\in R(t_i)} \boldsymbol{exp}(\beta X_r)}$ (5.3)
- Then finally obtaining the **Partial Likelihood** $\boldsymbol{L}(\beta)$ (5.4)

We shall assume that there are k different event of interest times such that there are no ties at the

$\boldsymbol{t_i}$, be the time of the happening of the **i**th event of interest. Where $\boldsymbol{i = 1, 2, \dots, k}$

Then the partial likelihood $\boldsymbol{L}(\beta)$ can be obtained as shown below

$$\boldsymbol{L(\beta)} = \prod_{i=1}^{k} \frac{h_i(t_i)}{\sum_{r\in R(t_i)} h_r(t_r)} \quad (5.5)$$

$$\boldsymbol{L(\beta)} = \prod_{i=1}^{k} \frac{exp(\beta X_i)}{\sum_{r\in R(t_i)} exp(\beta X_r)} \quad (5.6)$$

Solving the equation above enables us to obtain $\boldsymbol{exp}(\beta)$ and then β.

The steps to solving the equation are listed below.

- Obtain the partial likelihood $\boldsymbol{L(\beta)}$
- Obtain the $Log(\boldsymbol{L(\beta)})$, natural log (5.7)
- Then obtain the first derivative of the $Log(\boldsymbol{L(\beta)})$ with respect to β, $\frac{d\, Log\,(L(\beta))}{d\,\beta}$ (5.8)

With the below derivative

$\frac{d\,Log\,(L(\beta))}{d\,\beta} = \mathbf{0}$ **,** we can obtain $\boldsymbol{exp}(\beta)$ and then β

Hence we can estimate the B_l the **l**th regression parameter for the cox proportional hazards model.

To test the significance of the parameters obtained we can either use the Wald, Partial Likelihood ratio and Score Test.

We shall use the Wald to test significance of the parameters, using the hypotheses test below

$\boldsymbol{H_o}: \boldsymbol{\beta_l} = \mathbf{0}$

Versus

$\boldsymbol{H_1}: \boldsymbol{\beta_l} \neq \mathbf{0}$

Wald test statistic denoted by W, is obtained as below

$$W = \frac{\widehat{\boldsymbol{\beta}_l}}{se(\widehat{\boldsymbol{\beta}_l})}$$

$$W \sim \boldsymbol{H_o} \sim \boldsymbol{N}(\mathbf{0}, \mathbf{1})$$

Thus to conduct the hypothesis test for the hypotheses below

$\boldsymbol{H_o}: \boldsymbol{\beta_l} = \mathbf{0}$

Versus

$\boldsymbol{H_1}: \boldsymbol{\beta_l} \neq \mathbf{0}$

We obtain a critical value from the normal distribution tables at $\propto$ level of significance. If the **Test Statistic (**W**) is greater than the Critical Value** then reject$\boldsymbol{H_o}: \boldsymbol{\beta_l} = \mathbf{0}$

3.7.4 Interpretation of Coefficients

The interpretation of the exponentiated coefficients is the hazard/risk ratio of a one unit increase for the specific covariate. For example, consider a one unit increase in the covariate.

The coefficient by itself then has the interpretation of the logarithm of the hazard/risk ratio of a unit increase for the specific covariate. Note that the Cox Model does not estimate an intercept term. This is because the parameter is unidentifiable, that is, we are unable to estimate its value. Imagine that we added an intercept to this model. The model would then look like this:

$$\frac{h_0(t)\exp(\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\ldots.\beta_k(X_{k+1})+\ldots\ldots..+\beta pX_p)}{h_0(t)\exp(\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\ldots.\beta_k(X_k)+\ldots\ldots..+\beta pX_p)}=\exp(\beta_k)$$

The coefficient by itself then has the interpretation of the logarithm of the hazard/risk ratio of a unit increase for the specific covariate. Note that the Cox Model does not estimate an intercept term. This is because the parameter is unidentifiable, that is, we are unable to estimate its value. Imagine that we added an intercept to this model. The model would then look like this:

$$h(t|x)=\exp(\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\ldots.+\beta pX_p)=h_0(t)\exp(\beta_1X_1+\beta_2X_2+\beta_3X_3+\ldots.+\beta pX_p)$$

4.0 Results

4.1 Exploratory analyses - Results

Data was imported into the analysis software, stata, and analysis done.

The frequency of gender showed that 83.8% of the participants were male as shown in table 3 below

Table 3: Distribution based on gender

SEX	Distribution	Percent
Female	874	16.19
Male	4,523	83.81
Total	**5,397**	**100**

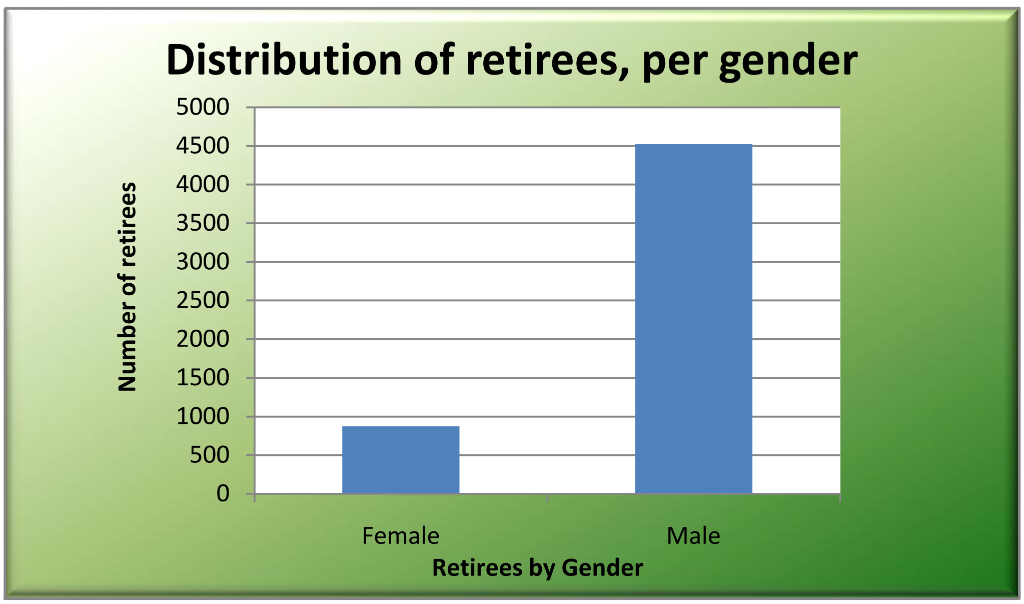

Figure 1: Pictorial representation of distribution per gender

Frequency based on categories according to monthly pension received suggest that majority of retirees earn between Kshs. 5000 to 10,000 while very few retirees earn more than Kshs. 100,000 as shown below:

Table 4: Distribution based on pension categories

Pension_Category	Number of retirees	Percentage
Less than 5000	989	18.32
5000 to 10000	2,020	37.43
10000 to 20000	1,696	31.42
20000 to 40000	585	10.84
40000 to 100000	89	1.65
More than 100000	18	0.33
Total	**5,397**	**100**

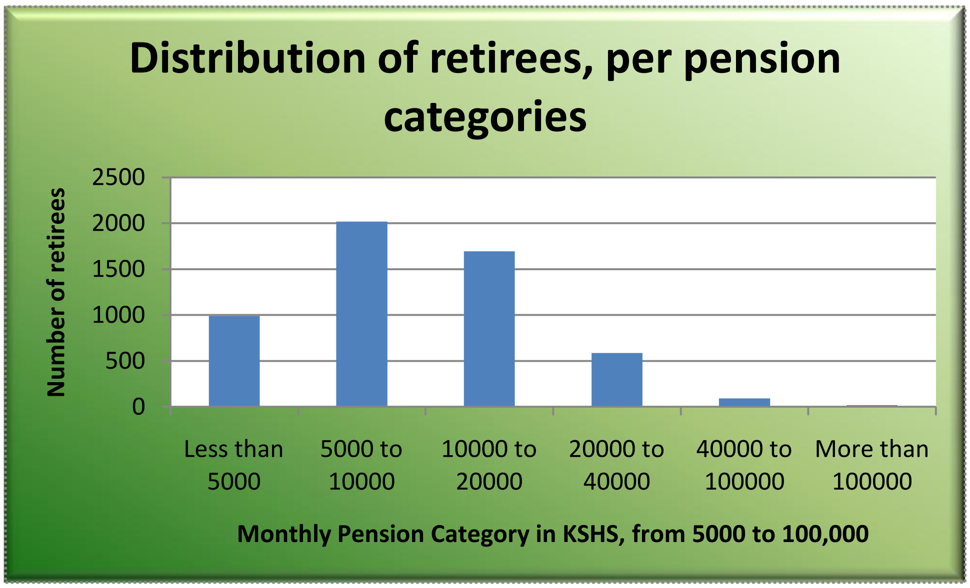

Figure 2: Pictorial representation of frequency per pension categories

As expected, most of the retirees were 55 years or less, since most of the retirees in this dataset retired before 2009, with only 278 retirees retiring after July 2009, when the retirement age in Kenya was changed to 60 years.

Table 5: Frequency based on age categories

Age_Category	Distribution	Percent
Less than 55	4,587	84.99
55-60	550	10.19
60-65	254	4.71
More than 65	6	0.11
Total	**5,397**	**100**

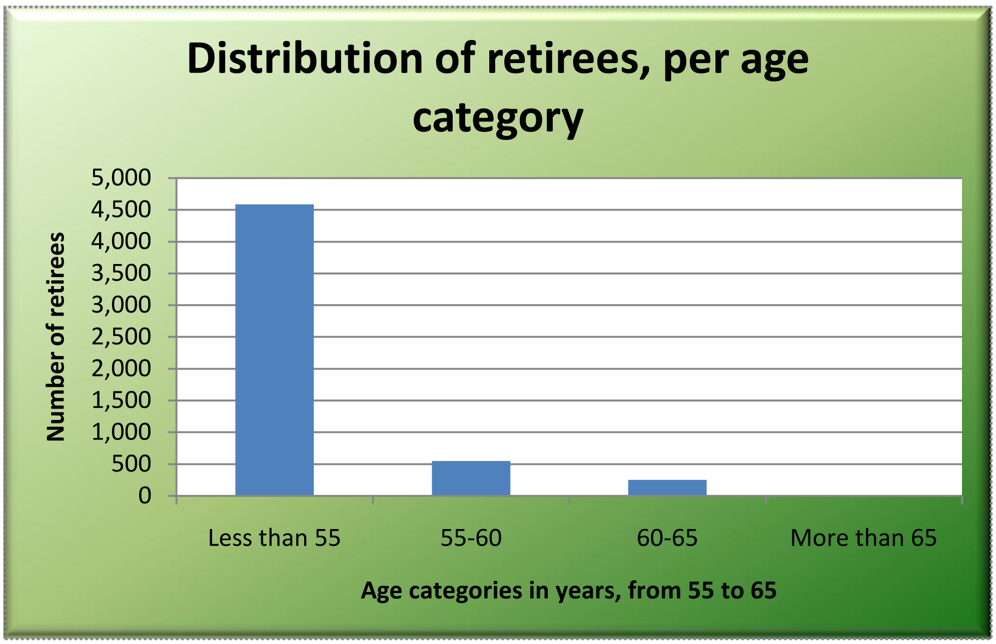

Figure 3: Pictorial representation of distribution per age categories

4.2 Life Table - by gender

Table 6: Males' life table, demonstrating life expectancy

Interval		Beg. Total	Deaths	Censored	Survival	Std. Error	[95% Conf. Int.]	
0	**1**	4523	2	33	0.9996	0.0003	0.9982	0.9999
1	**2**	4488	6	15	0.9982	0.0006	0.9964	0.9991
2	**3**	4467	13	8	0.9953	0.001	0.9928	0.9969
3	**4**	4446	22	165	0.9903	0.0015	0.9869	0.9928
4	**5**	4259	29	65	0.9835	0.0019	0.9793	0.9869
5	**6**	4165	40	189	0.9738	0.0024	0.9686	0.9782
6	**7**	3936	27	143	0.967	0.0028	0.9612	0.9720
7	**8**	3766	8	1405	0.9645	0.0029	0.9584	0.9697
8	**9**	2353	8	315	0.961	0.0031	0.9544	0.9667
9	**10**	2030	5	158	0.9585	0.0033	0.9515	0.9645
10	**11**	1867	7	134	0.9548	0.0036	0.9472	0.9613
11	**12**	1726	7	112	0.9508	0.0039	0.9426	0.9579
12	**13**	1607	5	56	0.9478	0.0041	0.9391	0.9552
13	**14**	1546	4	33	0.9453	0.0043	0.9363	0.9531
14	**15**	1509	9	72	0.9395	0.0047	0.9297	0.9480
15	**16**	1428	8	62	0.9342	0.005	0.9236	0.9433
16	**17**	1358	17	66	0.9222	0.0057	0.9102	0.9326
17	**18**	1275	16	60	0.9103	0.0064	0.8970	0.9220
18	**19**	1199	17	18	0.8973	0.007	0.8827	0.9102
19	**20**	1164	10	515	0.8874	0.0076	0.8716	0.9014
20	**21**	639	10	94	0.8724	0.0088	0.8540	0.8887
21	**22**	535	3	72	0.8672	0.0093	0.8478	0.8843
22	**23**	460	8	74	0.8508	0.0108	0.8283	0.8706
23	**24**	378	7	63	0.8336	0.0124	0.8077	0.8563
24	**25**	308	7	53	0.8129	0.0143	0.7829	0.8391
25	**26**	248	5	31	0.7954	0.016	0.7619	0.8247
26	**27**	212	6	43	0.7703	0.0185	0.7316	0.8042
27	**28**	163	3	23	0.7551	0.0201	0.7130	0.7919
28	**29**	137	6	22	0.7191	0.0239	0.6691	0.7629
29	**30**	109	2	21	0.7045	0.0256	0.6511	0.7513
30	**31**	86	3	19	0.6769	0.0291	0.6161	0.7302
31	**32**	64	2	7	0.6545	0.0322	0.5874	0.7134
32	**33**	55	2	17	0.6264	0.0364	0.5505	0.6930
33	**34**	36	0	9	0.6264	0.0364	0.5505	0.6930
34	**35**	27	1	1	0.6027	0.042	0.5152	0.6794
35	**36**	25	2	1	0.5535	0.051	0.4481	0.6467
36	**37**	22	1	1	0.5278	0.0547	0.4154	0.6282
37	**38**	20	0	3	0.5278	0.0547	0.4154	0.6282

38	**39**	17	1	2	0.4948	0.0604	0.3720	0.6061
39	**40**	14	0	5	0.4948	0.0604	0.3720	0.6061
40	**41**	9	0	1	0.4948	0.0604	0.3720	0.6061
41	**42**	8	0	1	0.4948	0.0604	0.3720	0.6061
42	**43**	7	0	3	0.4948	0.0604	0.3720	0.6061
43	**44**	4	0	1	0.4948	0.0604	0.3720	0.6061
44	**45**	3	0	1	0.4948	0.0604	0.3720	0.6061
46	**47**	2	1	1	0.1649	0.1915	0.0017	0.6006

Life table results implying that as at 15 years after retirement, 66.64% of the male retirees had died, having only 33.36% surviving.

Table 7: Females' life table, demonstrating life expectancy

Interval		**Beg. Total**	**Deaths**	**Censored**	**Survival**	**Std. Error**	**[95% Conf. Int.]**	
0	**1**	874	0	12	1	0	.	.
1	**2**	862	1	14	0.9988	0.0012	0.9917	0.9998
2	**3**	847	2	3	0.9965	0.002	0.9891	0.9989
3	**4**	842	3	35	0.9928	0.0029	0.9841	0.9968
4	**5**	804	4	19	0.9878	0.0038	0.9775	0.9934
5	**6**	781	2	59	0.9852	0.0042	0.9741	0.9916
6	**7**	720	3	51	0.981	0.0049	0.9686	0.9885
7	**8**	666	0	372	0.981	0.0049	0.9686	0.9885
8	**9**	294	0	86	0.981	0.0049	0.9686	0.9885
9	**10**	208	0	29	0.981	0.0049	0.9686	0.9885
10	**11**	179	0	36	0.981	0.0049	0.9686	0.9885
11	**12**	143	0	16	0.981	0.0049	0.9686	0.9885
12	**13**	127	0	8	0.981	0.0049	0.9686	0.9885
13	**14**	119	0	8	0.981	0.0049	0.9686	0.9885
14	**15**	111	0	15	0.981	0.0049	0.9686	0.9885
15	**16**	96	0	7	0.981	0.0049	0.9686	0.9885
16	**17**	89	1	5	0.9696	0.0123	0.9334	0.9863
17	**18**	83	0	8	0.9696	0.0123	0.9334	0.9863
19	**20**	75	0	56	0.9696	0.0123	0.9334	0.9863
20	**21**	19	0	5	0.9696	0.0123	0.9334	0.9863
21	**22**	14	0	5	0.9696	0.0123	0.9334	0.9863
22	**23**	9	0	3	0.9696	0.0123	0.9334	0.9863
23	**24**	6	0	1	0.9696	0.0123	0.9334	0.9863

24	**25**	5	0	1	0.9696	0.0123	0.9334	0.9863
25	**26**	4	0	1	0.9696	0.0123	0.9334	0.9863
27	**28**	3	1	0	0.6464	0.264	0.0650	0.9327
33	**34**	2	1	0	0.3232	0.2639	0.0095	0.7605
35	**36**	1	1	0	0	.	.	.

Life table analysis results suggest that 87.3% of female retirees would have died 15 years after retirement and estimated life expectancy for males and females who retired at 55 years of age is 12 and 8 years respectively.

4.3 Kaplan-Meir Survival Curves

The survival curve without stratification is as shown in the KM curve below:

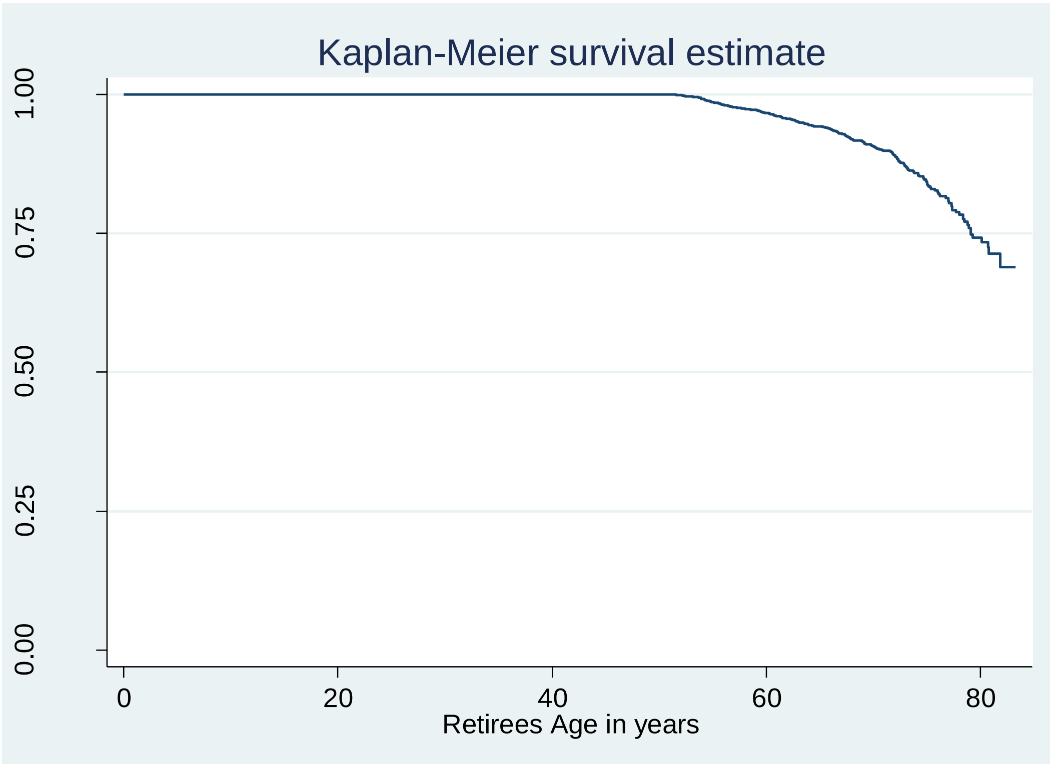

Figure 4: Survival curves for the retirees' data

As mentioned before, the survival curve remains at a plateau until an event, in this case death after retirement is observed. The origin of the data was specified to be the date of birth, which then means that from this age up to the first retirement, at age 50, we expect a plateau, as observed above. After the ages of retirement are reached, the curve continues in the same manner, as a plateau, since no death was observed, having the first death at age, 54 years.

Comparing survival curves by gender

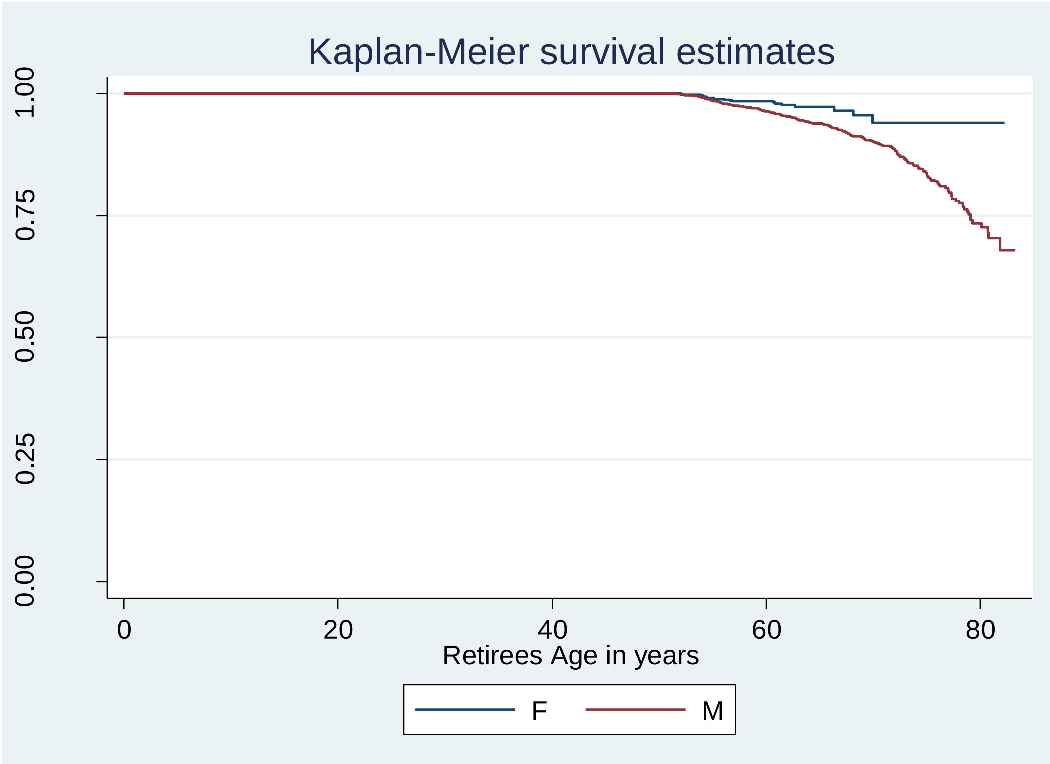

Figure 5: Survival curves comparing males and females' survival curves

The curve shows a significant difference between the survival times between men and women, with women having higher survival rates as compared to men.

Comparing survival curves by Pension Category

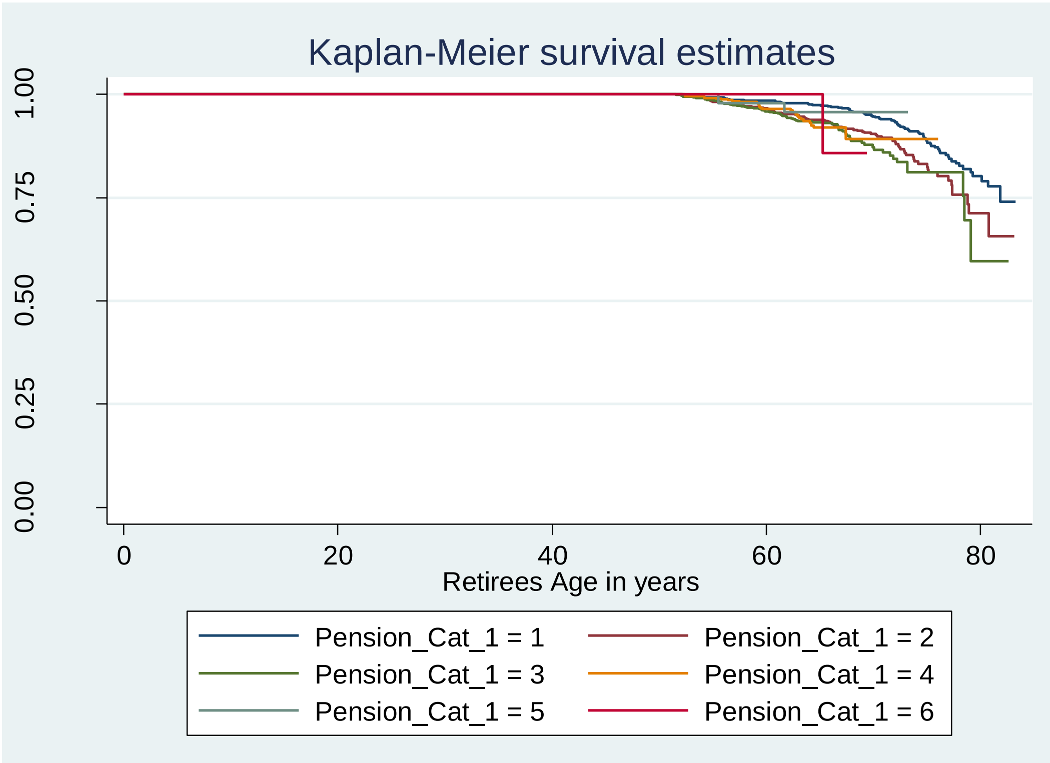

Figure 6: Survival curves comparing pension categories

The curve shows a significant difference between the categories of pension received monthly, with the categories considered as:

<5000 coded as 1

>5000 and <=10000 coded as 2

>10000 and <=20000 coded as 3

>20000 and <=40000 coded as 4

>40000 and <=100000 coded as 5

>100000 coded as 6

4.4 Inference – Log rank test

Log-rank test for equality of survivor functions - variable Sex

SEX	Events observed	Events expected
F	16	32.92
M	297	280.08
Total	313	313
chi2(1)		**9.99**
Pr>chi2		**0.0016**

Table 8: Equality of survivor function, by sex

At 95% confidence interval, α=0.05, thereby making the test significant, since the p-value is 0.0016, much less than 0.

Since the test is significant, the test statistic,

$$Log - rank\ statistic\ for\ two\ groups = \frac{\{\sum_{t_j}[d_{ij} - \frac{d_j n_{ij}}{n_j}]\}2}{\sum_{t_j}\frac{n_{1j}n_{2j}d_j(n_j-d_j)}{n_j^2(n_j-1)}} = 9.99$$

from the tests done, this means that the Hypothesis:

$\boldsymbol{H_o: S_{(a)} = S_{(b)}}$

Versus

$\boldsymbol{H_1: S_{(a)} \neq S_{(b)}}$

where, we let $S_a(t)$ be the survival function for males and $S_b(t)$ be the survival function for females which is compared to the X^2,1 = 3.84, (1 degree of freedom, since the groups considered are only 2).

Rejection criteria: If the calculated value greater than X^2, 1, we reject Ho and conclude that $S_a(t) \neq S_b(t)$, thus concluding that the males and females have different survival times.

Log-rank test for equality of survivor functions - variable Pension Category

Pension Category	Events observed	Events expected
1	73	107.46
2	120	108.93
3	91	71.6
4	26	21.45
5	2	3.01
6	1	0.55
Total	**313**	**313**
chi2(5)	**=**	**22.39**
Pr>chi2	**=**	**0.0004**

Table 9: Equality of survivor function, by pension category

At 95% confidence interval, α=0.05, since the p-value less than 0.001, much less than 0.05, we conclude that the test is significant. Since the test is significant, the test statistic:

$$Log - rank\ statistic\ for\ two\ groups = \frac{\{\sum_{t_j}[d_{ij} - \frac{d_j n_{ij}}{n_j}]\}2}{\sum_{t_j}\frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}} = 22.39$$

from the tests done, Hypothesis: H_0: $S_a(t) = S_b(t)$ versus $S_a(t) \neq S_b(t)$

Where, we let $S_a(t)$ be the survival function for males and $S_b(t)$ be the survival function for females which is compared to the $X^2,1 = 3.84$, (1 degree of freedom, since the groups considered are only 2).

Rejection criteria: If the calculated value greater than $X^2,1$, we reject Ho and conclude that $S_a(t) \neq S_b(t)$, thus concluding that the 6 categories of monthly pension received by the retirees have different survival times, significantly different.

4.5 Inference - Cox-proportional regression model

Using the model,

$$h(t|x) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$$

We analyse the covariates, X, to determine their significance in determining the time to death, which is our dependent variable.

X1 is sex

X2 is age-at-retirement

X3 is pension (to measure the socio-economic status)

Results:

Numbers of subjects observed were 5250, with 313 experiencing the event, death after retirement. The p-value is less than 0.001

Covariates	Haz. Ratio	Std. Err.	z	P>z	[95% Conf.	Interval]
SEX_1	2.10	0.54	2.86	0.004	1.26	3.49
Ageatretirement	0.96	0.02	-1.75	0.08	0.91	1.01
Pension_Cat_1	1.26	0.07	3.83	0	1.12	1.41

Table 10: Cox regression analysis test results

4.5.1 Significance of test:

At 95% confidence interval, α=0.05, thereby making the model being tested significant, since the p-value is less 0.001, much less than 0.05.

4.5.2 Interpreting the coefficients

Sex: X1,

H0:β_0=0 vs H1: $\beta_0 \neq 0$

At 95% confidence interval, =0.05,since the p-value is 0.004, much less than 0.05, we reject H0 above, thereby making the test for variable Sex, significant since $\beta_0 \neq 0$Considering that Sex is a categorical variable, with female, coded as 0, considered as the reference group, the hazard ratio value, 2.10, will be interpreted as:

Males are 2.10 more likely to experience death after retirement as compared to females, Meaning, females live/survive 2.10 times longer than males.

Age-at-retirement: X2,

H0: β_1=0 vs H1: $\beta_1 \neq 0$

At 95% confidence interval, α=0.05, since the p-value is 0.090, which is greater than 0.05, we fail to reject H0 above, thereby making the test for variable Age-at-retirement not significant, since β_1=0.

However, we still consider the result, since Age-at-retirement is a continuous variable; results will be measured in regards to a unit increase of the value.

Hazard ratio, 0.96, being less than 1, we subtract the value from 1, to get 0.04. Result showing, that for every unit increase in the age at retirement, the chances of experiencing death are reduced by 4%.

Pension Category (Amount paid as monthly pension, alluding to socio-economic status):X3,

H0: β_2=0 vs H1: $\beta_2 \neq 0$

At 95% confidence interval, α=0.05,since the p-value is less than 0.001, much less than 0.05, we reject H0 above, thereby making the test for variable Pension, significant, since $\beta_2 \neq 0$.

Considering that Pension Category is a categorical variable, with categories:

<5000 coded as 1

>5000 and <=10000 coded as 2

>10000 and <=20000 coded as 3

>20000 and <=40000 coded as 4

>40000 and <=100000 coded as 5

>100000 coded as 6

with <5000, code 1, considered as the reference group.

The hazard ratio for this variable is, 1.255072, being greater than 1, we subtract 1 from the value, to get a 25.5%. The result showing that the group receiving less than 5,000 KSHS monthly is 25.5% more likely to experience the death, as compared to the other groups receiving more money than that.

5.0 Discussion

The study has determined life expectancy of male and female retirees, with the age at retirement based on the difference between the date of birth and date of retirement. The results show that the life expectancy of the male retirees is 12 years while that of their female counterparts is 8 years. On running the regression models, using the cox proportional hazard models, we observed that male retirees were 2.10 times more likely to die after retirement as compared to their female counterparts. However, there is clear difference between survival times for males and females, with females having longer survival times, specifically, 2.18 times longer as compared to males, even though, Shan, [4], found that the difference in survival could not be attributed to the gender.

The age-at-retirement, though the test was not significant, does determine the likelihood for survival, up to 4%, with reduction in the risk of death, or otherwise stated, increase in the probability of survival, up to 4% with every unit increase. This means that the longer one delays retirement, the more their chances of living longer, up to 4%, for every year. Hilke, [3], attributed this increase in probability of survival to the health status of the retiree. He mentions that most probably people, who retire early, retire because they have to, due to their ill health. This therefore leaves a question, should more studies be done, with the health status of the retiree included in the model, to assess whether the health status is indeed a confounder?

There was a significant difference in the different probability of surviving among the retirees, based on the amount of monthly pension received, with the group receiving the least, less than 5000 KSHS, being 25.5% more likely to experience death as compared to the groups earning more pensions, Haynes, [11] mentioned that the retirees of lower status were more likely to die compared to those of higher status, those earning more money. Therefore, we conclude that the amount of money paid monthly, as pension, is a significant factor on the time to death after retirement, probably because it has a direct implication the living standards of individual.

In Kenya, it has been observed that after retirement, most people move to the rural areas, where life is very different, much slower, yet they were used to the urban areas, where life was extremely fast. Most retirees have been reported to become extremely inactive, thus reducing their days on earth. Their motivation has also been seen to reduce, since there is not much activity around them, this has also been seen to be a determinant to the time to death after retirement.

6.0 Conclusion and recommendations

This study concluded that early retirement does indeed increase the chances of early death, though the study also suggests that more information on the health status of the early retirees should be used as a covariate during such a study, to eliminate the possibility that the early retirees were already sick, thus their health status rather that their age cause the early death. Retirees should be advised to plan for their retirement since those retirees who earned less were 25% more likely to die than their counterparts who earned more. The standards of living after retirement can be boosted by a better savings plan that ensures that the life after retirement is catered for therefore employers should educate their employees on planning for this.

Since females have been seen to live longer than males, it would be interesting to see the effect of marital status on time to death for male retirees.

Health status should be associated with the age-at-retirement to adjust for the covariate of ill-health as a determinant to early death, after early retirement.

Profession of the retirees, their residence before and after retirement should also be considered in other studies, to determine if this has a direct effect on mortality. Pension schemes and life annuity programs in insurance companies should be sensitized on the importance of comprehensive data, to be collected before enrolling individuals since lack of these variables was clearly a limitation to this study.

Employers should sensitize their employees on the importance of planning for retirement, since standards of living (socio-economic status) after retirement do have a direct impact on years lived after retirement.

7.0 References

[1]Sing Lin, Ph.D, March 2002. Optimum Strategies for Creativity and Longevity http://faculty.kfupm.edu.sa/COE/gutub/English_Misc/Retire1.htm

[2] Litwin, H. (2007). Does early retirement lead to longer life? Ageing & Society, 27, 739—754

[3] Hilke Brockmanna, Rolf Mu¨ller b, Uwe Helmert b, 2009, Time to retire -- Time to die? A prospective cohort study of the effects of early retirement on long-term survival, Social Science & Medicine 69 (2009) 160—164

[4] Shan P Tsai,Judy K Wendt,Robin P Donnelly,Geert de Jong, Farah S Ahmed, 2005, Age at retirement and long term survival of an industrial population: prospective cohort study. BMJ 2005;331:995, http://www.bmj.com/content/331/7523/995

[5] Mark Stibich, Ph.D, 2009. Is retirement good for you?

[6] Angelo Tinazzi, SENDO-Tech S.r.l., Martin Scott, Numerus Ltd, Tuebingen, Anna Compagnoni, SENDO-Tech S.r.l.,2008, A gentle introduction to survival analysis

[7] Nihal Ata, M. Tekin Sozer, 2007, Cox Regreesion Models With Nonproportional hazards applied to Lung Cancer survival data. Volume 36 (2) (2007), 157-157

[8]Thomas E. Nesthus, Stephen J. H. Veronneau, Satya P. Sangal,Robert O. Besco, 1995, A Longevity and Survival Analysis for a Cohort of Retired Airline Pilots, DOT/FAA/AM-95/5

[9]Tyler Smith, Kaplan Meier And Cox Proportional Hazards Modeling:
Hands On Survival Analysis

[10] ATCHLEY, R. C. (1994). Retirement: Leaving the World of Work. Sage Journals .

[11] Haynes, S. G. (1978). Survival After Early and Normal Retirement. The Journal of Gerentology

[12] Mein, G. (2000). Predictors of early retirement in British civil servants. Oxford Journals .

[13] Mwangi, F. (2013). Analysis of Longevity Risk and its effect on Retirement benefits.

[14] Global action on aging, http://www.globalaging.org/pension/world/archive09.htm

[15] Michael E. Leonetti, 2009. Retirement Planning Myths